Limitations of the Parsonnet score for measuring risk stratified mortality in the north west of England

K Wynne-Jones, M Jackson, G Grotte, B Bridgewater and North West Regional Cardiac Surgery Audit Steering Group

Heart 2000;84;71-78
doi:10.1136/heart.84.1.71

Limitations of the Parsonnet score for measuring risk stratified mortality in the north west of England

K Wynne-Jones, M Jackson, G Grotte, B Bridgewater, on behalf of the North West Regional Cardiac Surgery Audit Steering Group

Abstract

Objective—To study the use of the Parsonnet score to predict mortality following adult cardiac surgery.

Design—Prospective study.

Setting—All centres performing adult cardiac surgery in the north west of England.

Subjects—8210 patients undergoing surgery between April 1997 and March 1999.

Main outcome measures—Risk factors and in-hospital mortality were recorded according to agreed definitions. Ten per cent of cases from each centre were selected at random for validation. A Parsonnet score was derived for each patient and its predictive ability was studied.

Results—Data collection was complete. The operative mortality was 3.5% (95% confidence interval 3.1% to 3.9%), ranging from 2.7% to 3.8% across the centres. On validation, the incidence of discrepancies ranged from 0% to 13% for the different risk factors. The predictive ability of the Parsonnet score measured by area under the receiver operating characteristic curve was 0.74. The mean Parsonnet score for the region was 7.0, giving an observed to expected mortality ratio of 0.51 (range 0.4 to 0.64 across the centres). A new predictive model was derived from the data by multivariate analysis which includes nine objective risk factors, all with a significant association with mortality, which highlights some of the deficits of the Parsonnet score.

Conclusions—Risk stratified mortality data were collected on 100% of patients undergoing adult cardiac surgery in two years within a defined geographical region and were used to set an audit standard. Problems with the Parsonnet score of subjectivity, inclusion of many items not associated with mortality, and the overprediction of mortality have been highlighted.

(*Heart* 2000;**84**:71–78)

Keywords: risk stratification; cardiac surgery; Parsonnet score; audit

Departments of Cardiothoracic Surgery and Clinical Audit, Wythenshawe Hospital, Southmoor Road, Manchester M23 9LT, UK
K Wynne-Jones
B Bridgewater

Departments of Cardiothoracic Surgery and Clinical Audit, The Cardiothoracic Centre–Liverpool, Thomas Drive, Liverpool L14 3PE, UK
M Jackson

Department of Cardiothoracic Surgery, Manchester Royal Infirmary, Oxford Road, Manchester M13 9WL, UK
G J Grotte

Correspondence to:
Dr Bridgewater email:
ben@bridgewaterb.fsnet.co.uk

Accepted 8 February 2000

Measuring the outcome of health interventions has become a national priority,[1] both to demonstrate that a particular treatment is usually effective and to show that the intervention itself has been performed to a satisfactory standard. In no branch of medicine has this second area been discussed more widely than in cardiac surgery.[2 3] For many years the Society of Cardiothoracic Surgeons of Great Britain and Ireland has collected anonymised mortality statistics for cardiac and thoracic surgical procedures, which are pooled and published as national figures in an annual register.[4] This has enabled trends to be followed and institutions or individual surgeons to compare their results with these standards. Recently the Society of Cardiothoracic Surgeons has asked all units undertaking adult cardiac surgery to submit mortality rates for each consultant for first time coronary artery surgery in an effort to "police" the specialty. However, it is well known that outcome following cardiac surgery is dependent on various preoperative patient characteristics[5]; such data may therefore be misleading as there are probably quite notable local and regional variations in the incidence of these risk factors.[6]

Various models have been developed to predict mortality from preoperative patient characteristics, and the first to become popular was the Parsonnet risk stratification system,[7] which was derived in the USA in the 1980s. It has been shown to be applicable to British cardiac surgical practice to some extent,[8] and in a recent multicentre study it was found to be the best available predictor of coronary artery surgery mortality in the UK.[6] However, it has been criticised for the nature of its statistical derivation,[6 9 10] in that it systematically overestimates mortality, particularly for high risk patients, and because its scoring system is quite subjective, again especially in the high risk group. It also omits many surgeons' "favourite" risk factors, such as the number of coronary vessels diseased, urgency of operation, and the presence of chronic obstructive pulmonary disease.[11]

In spite of the focus on results in cardiac surgery in recent years, there are still no accepted risk stratified mortality standards for the UK. We have collected preoperative patient characteristics and operative mortality on all patients undergoing adult cardiac surgery in a defined geographical area in the UK, validated the dataset, and studied the use of the Parsonnet score to predict mortality in this population.

Methods

Data were collected on all patients undergoing adult cardiac surgery between 1 April 1997 and 31 March 1999 at the four centres in the north west of England. Patients undergoing aortic

surgery, surgery for cardiac trauma or tumour, adult congenital surgery, or closed cardiac surgical procedures were excluded from the analysis (these patients are categorised as "miscellaneous" for the UK cardiac surgical register[4]). Patients having surgery for complications of ischaemic heart disease such as left ventricular aneurysm repair, ventricular septal defect surgery, and ischaemic mitral valve disease were also excluded (these patients are categorised as "miscellaneous ischaemic heart disease" for the cardiac surgical register[4]), as were those undergoing thoracic transplantation. Slightly different techniques of data collection were used in the four centres. The definitions used for the various risk factors are shown in the appendix. These definitions are similar to those given in the UK Society of Cardiothoracic Surgeons "minimum dataset" but were modified slightly by consensus between the participating centres before starting the study. Data were returned to a central base, where a Parsonnet score was calculated for each patient.

ROYAL VICTORIA INFIRMARY, BLACKPOOL AND THE CARDIOTHORACIC CENTRE, LIVERPOOL

Patient data were collected by the patient administration and tracking system (PATS, Dendrite Clinical Systems, London, UK), whereby risk factors, operation data, and outcome data are recorded on a structured form in the patient's notes. Preoperative medical characteristics were recorded by a junior doctor, and operation data detailed by the surgeons and perfusionists. The data were transcribed onto the PATS software by a data entry clerk at a later date.

MANCHESTER ROYAL INFIRMARY

All data were entered directly onto a database by medical personnel using customised software (Omnis 7, Omnis Software Inc, Foster City, California, USA). The preoperative medical characteristics were entered by preregistration house surgeons, the operation data by the operating surgeon and the perfusionist, and the outcome data by medical personnel at the time of discharge or death.

WYTHENSHAWE HOSPITAL

Risk factor and operative data were entered onto the database (Foxpro, Microsoft, Redmond, Washington, USA) by the operating surgeon at the time of surgery, using data collected by a junior doctor at the time of admission on a structured patient clerking sheet. Outcome data were entered on the database at the time of discharge.

RISK PREDICTION ALGORITHMS

The Parsonnet score was used as a risk prediction algorithm.[7] The model allocates additive predicted mortality percentage points for 14 patient risk factors to give a "Parsonnet score",[7] which is indicative of the per cent mortality for each patient (table 1). A score of 10 points was awarded for patients in a catastrophic state, rather than allowing between 10 and 50 as suggested in the original Parsonnet system, in order to decrease subjectivity in the higher risk group.[6]

OPERATIVE MORTALITY

Mortality was recorded from patient and hospital records. Each centre had a dedicated person to collect these data, and detailing of mortality was exhaustive. Operative mortality was defined as "death within the same hospital admission as operation, regardless of cause." All patients transferred from the base hospital to another hospital were followed up to confirm their status at discharge.

DATA VALIDATION

Data from all the centres were returned to a central base for analysis. They were checked for missing or obviously erroneous records, which were returned to the relevant centre for completion or correction. Once a complete dataset was obtained for the first three months, 10% of cases were selected at random from each of the centres and the patient records were studied by a surgically trained observer to look for discrepancies. New Parsonnet scores were then derived, based on the data obtained through the validation checks. The incidence of discrepancies between submitted and validated data for the risk factors was tabulated for the region. The derived Parsonnet scores from submitted and validated data were compared by a two tailed paired Student's *t* test.

STATISTICAL ANALYSIS

The number of cases in each centre and the average number of cases per consultant in each centre were determined and compared; the types of operation performed in the region and in each centre were compared; and the incidence of the risk factors in the region and in each centre were tabulated and compared. The mortalities for each centre and for each type of operation were determined and compared with regional figures and with UK national figures.[4] All the above comparisons were performed using χ^2 tests.

Table 1 The Parsonnet score

Risk factor	*Score*
Female	1
Obesity, > 1.5 times ideal weight	3
Diabetes	3
Hypertension	3
Ejection fraction (%)	
> 50	0
30–49	2
< 30	4
Age (years)	
71–74	7
75–79	12
⩾ 80	20
First reoperation	5
Second reoperation	10
Preoperative IABP	2
Emergency from procedures laboratory	10
Dialysis dependent	10
Catastrophic states	10
Valve surgery	
Mitral	5
PAP > 60 mm Hg	3
Aortic	5
Gradient > 120 mm Hg	2
CABG with valve	2

CABG, coronary artery bypass graft; IABP, intra-aortic balloon pump; PAP, pulmonary artery systolic pressure.

The ability of the Parsonnet score to predict observed mortality was determined by measuring the area under the receiver operating characteristic (ROC) curve.[12–15] The ROC curve is a plot of sensitivity against 1−specificity and is generally regarded to be a good summary measure of the predictive ability of this type of algorithm. An area of 1 suggests a perfect predictor, 0.5 suggests a predictor that is no better than chance alone, and scores of between 0.7 and 0.9 are generally regarded as useful.

Predicted mortalities between the centres were compared by analysis of variance (ANOVA). A significant variation among the means was further analysed using Duncan's multiple comparisons tests. The ratio of observed mortality to that predicted by the Parsonnet score was determined and used as an index of operative performance. Although data were not normally distributed, it has been accepted that the mean should be used to compare expected mortality and to calculate the observed to expected mortality ratios[6 16 17].

The association between risk factors and mortality was determined by multivariate analysis of the first 18 months dataset, comprising 6246 patients, using stepwise logistic regression and "bootstrapping". Variables were accepted into the model on the basis of a significant association with mortality ($p < 0.05$) and the 95% confidence intervals (CI) of the odds ratios not spanning 1. Bootstrapping[18] is a process of random sampling with replacement from the original dataset; therefore some observations may be included in a sample 10 times and some may not be included at all. Drawing 1000 samples of 6246 observations gives 1000 estimates of the odds ratios, and the distribution of these estimates provides a better overall estimate of the standard error and the confidence intervals of the odds ratio. The model was validated using ROC curve analysis on an independent dataset of 1958 patients.

Cusum curves[19] were plotted for each of the centres, with the number of patients plotted along the x axis and cumulative mortality along the y axis, with 95% CI around each point. These curves were compared with cumulative predicted mortality from the population as determined by the Parsonnet score. As the Parsonnet score significantly overpredicts observed risk, we adjusted it by the observed to expected mortality ratio for the whole region.

Results

In all, 8633 patients underwent adult cardiac surgery in the northwest region between 1 April 1997 and 31 March 1999. On the basis of operation performed, 423 patients were excluded from further analysis. These patients had a mortality of 12%.

RISK FACTOR VALIDATION

In all, 101 cases were validated. The incidence of discrepancies between submitted and validated data for the risk factors for the region as a whole is shown in table 2. The following risk factors had an incidence of discrepancy greater than 10%: previous Q wave myocardial infarction (13%), hypercholesterolaemia (12%), and hypertension (11%). The mean Parsonnet scores, as determined by submitted and validated data, were 7.4 and 5.6, respectively ($p = 0.09$). Nineteen per cent of all cases validated had differences in Parsonnet score between submitted and validated data. The maximum difference was 12 Parsonnet points.

ACTIVITY

After exclusions, the number of cases performed at centres 1, 2, 3, and 4 were 1443, 2880, 1714, and 2173, respectively. There were significant differences in the average numbers of cases performed by independent operators between the four centres, and they ranged from 360 to 429 cases over the two years ($p = 0.02$). However, some of the surgeons in the region are dedicated cardiac surgeons and some are combined cardiothoracic surgeons, and the proportion of each at the hospitals and the relative thoracic workloads at each hospital are different, making direct comparisons of cardiac activity somewhat difficult. Coronary artery surgery was responsible for 78% of all activity, followed by aortic valve replacement (8%), combined aortic valve replacement and coronary artery surgery (6%), and mitral valve replacement (4%). There were significant differences in case mix across the four centres ($p < 0.01$ for coronary artery bypass graft surgery, aortic valve replacement, and mitral valve replacement).

INCIDENCE OF RISK FACTORS

There was a significant difference in the incidence of most of the risk factors across the region (table 3).

MORTALITY

The overall mortality for the region was 286 of 8210 cases (3.5%, 95% CI 3.1% to 3.9%), similar to national annual figures for the UK in 1997/8[4]: 1179 of 33 615 cases (3.5%, 95% CI 3.3% to 3.7%) ($p = 0.1$). There were no significant differences in crude mortality across centres ($p = 0.2$), as shown in table 4. The national mortality, regional mortality, and mortality for each centre for each type of

Table 2 Per cent discrepancies for the region (101 records)

Field	*Region*
Previous Q wave myocardial infarction	12.9
Hypercholesterolaemia	11.9
Hypertension	10.9
Symptom status	9.9
Smoker	9.4
Ejection fraction	7.9
Disease	6.9
Previous cardiological intervention	5.0
Weight	5.0
Respiratory failure	4.0
Priority	4.0
Cerebrovascular disease	3.0
Peripheral vascular disease	3.0
Diabetes	2.0
Intravenous nitrates	2.0
Renal failure	1.0
Sex	0
Ventilated	0
Previous operations	0
Height	0
Operation type	0
Age	0

Table 3 Per cent incidence of risk factors in each centre with 95% CI for coronary artery bypass graft only (4929 patients)

Field	Centre 1 (1159 patients)	Centre 2 (2173 patients)	Centre 3 (1403 patients)	Centre 4 (1693 patients)	p Value (χ^2)*
Current smoker	7 (5.3 to 8.3)	14 (12.8 to 15.8)	13 (11.2 to 14.8)	9 (7.3 to 10.0)	<0.0001
Ejection fraction < 50%	34 (31.4 to 36.9)	45 (42.4 to 46.6)	35 (32.6 to 37.6)	24 (22.0 to 26.1)	<0.0001
Hypercholesterolaemia	61 (58.1 to 63.8)	67 (65.4 to 69.4)	77 (74.2 to 78.7)	67 (65.0 to 69.0)	<0.0001
Intravenous nitrates	13 (10.9 to 14.9)	8 (6.9 to 9.2)	10 (8.2 to 11.4)	4 (3.3 to 5.3)	<0.0001
Non-elective operation	33 (30.2 to 35.7)	20 (18.6 to 22.1)	27 (23.9 to 28.6)	14 (12.8 to 16.2)	<0.0001
Redo operation	4 (2.7 to 4.9)	3 (2.1 to 3.6)	7 (5.7 to 8.5)	4 (2.9 to 4.8)	<0.0001
Respiratory failure	13 (10.9 to 14.9)	29 (27.2 to 31.1)	9 (7.8 to 10.9)	11 (9.9 to 13.0)	<0.0001
Unstable symptoms	33 (29.9 to 35.3)	18 (16.8 to 20.1)	27 (25.0 to 29.7)	17 (15.3 to 18.9)	<0.0001
Peripheral vascular disease	15 (13.1 to 17.3)	12 (10.7 to 13.5)	11 (9.1 to 12.4)	8 (6.5 to 9.1)	<0.0001
Age > 70 years	19 (17.3 to 21.9)	21 (19.1 to 22.5)	16 (13.7 to 17.6)	15 (13.2 to 16.7)	<0.0001
Previous cardiological intervention	10 (8.8 to 12.4)	7 (6.0 to 8.2)	7 (5.4 to 8.1)	5 (4.1 to 6.2)	<0.0001
Female sex	20 (17.7 to 22.4)	19 (17.7 to 21.0)	24 (22.0 to 26.6)	20 (18.3 to 22.1)	0.003
Renal failure	1 (0.8 to 2.3)	1 (0.9 to 2.0)	2 (1.4 to 2.9)	3 (2.2 to 3.8)	0.003
Cardiogenic shock	1 (0.7 to 2.1)	0.4 (0.2 to 0.8)	0.6 (0.3 to 1.2)	0.1 (0 to 0.5)	0.008
Cerebrovascular disease	9 (7.5 to 10.9)	9 (7.4 to 9.8)	7 (5.7 to 8.5)	7 (5.5 to 7.9)	0.03
Hypertension	43 (40.5 to 46.3)	45 (42.6 to 46.8)	40 (37.7 to 42.9)	44 (41.3 to 46.1)	0.07
Ventilated	0.3 (0.1 to 1.0)	0.2 (0.1 to 0.6)	0.1 (0 to 0.5)	0 (0 to 0.2)	0.08
Previous Q wave MI	51 (47.8 to 53.6)	49 (46.4 to 50.7)	47 (44.3 to 49.6)	50 (47.3 to 52.1)	0.2
Diabetes	15 (12.6 to 16.8)	15 (13.6 to 16.6)	16 (13.8 to 17.6)	14 (12.0 to 15.3)	0.4
Failed intervention	1 (0.4 to 1.6)	0.6 (0.3 to 1.0)	0.8 (0.4 to 1.4)	0.4 (0.2 to 0.9)	0.6
BMI > 35 (kg/m^2)	5 (3.9 to 6.6)	4 (3.5 to 5.2)	4 (3.2 to 5.3)	5 (3.6 to 5.6)	0.7

*The p value refers to a comparison of the incidence of risk factors across the four centres.
BMI, body mass index; MI, myocardial infarction.

operation are shown in table 5. There was no significant difference in mortality between the four centres for any of the operation types.

PREDICTIVE ABILITY OF PARSONNET SCORE

The area under the ROC curve for the Parsonnet score for the region was 0.74, suggesting a model of useful predictive ability.

COMPARISON OF PREDICTED MORTALITY BETWEEN CENTRES

The mean Parsonnet score for the region was 6.9 (median 5, range 0 to 50). The mean scores for the individual centres are shown in table 4. These differences were significant ($p < 0.0001$), with centre 4 having a lower predicted risk than the other three centres, and centres 1 and 3 having a lower predicted risk than centre 2.

The overall mortality for the region was 3.5% and the mean Parsonnet score was 6.9, giving an observed to expected ratio for the region of 0.51. This correction factor has been applied for the remainder of the analysis. The corrected observed to expected mortality ratios for the centres are shown in table 4 and range from 0.85 to 1.1.

Table 4 Observed mortality, expected mortality, and observed to expected (O:E) mortality ratios by centre

	Mortality (95% CI)	Mean Parsonnet score	SD	Median Parsonnet score	Range	O:E ratio*
Centre 1	3.6 (2.7 to 4.7)	7.0	6.8	5	0 to 39	1.0
Centre 2	3.8 (3.2 to 4.6)	7.6	7.0	6	0 to 50	0.98
Centre 3	2.7 (2.0 to 3.7)	6.6	6.6	5	0 to 42	0.85
Centre 4	3.5 (2.8 to 4.4)	6.2	6.3	4	0 to 34	1.13

*O:E ratio has been calculated working to 0.51 of the Parsonnet score.

PREDICTIVE MODELLING

Multivariate analysis showed that the factors significantly associated with increased mortality were age, renal failure, ejection fraction < 50%, hypertension, non-elective surgery, redo surgery, acute preoperative state (in cardiogenic shock, or intra-aortic balloon pump inserted preoperatively not for prophylactic purposes), preoperative ventilation, double valve surgery, and combined coronary artery bypass graft and mitral valve surgery (table 6).

The odds ratios in table 6 should be compared with the weightings for those items in the Parsonnet score that are also given in the table. In addition, the table shows the discrepancy rates in the significant fields, taken from the validation checks described previously. With the exception of hypertension, the only fields with a moderate discrepancy rate were ejection fraction, with an 8% incidence, and non-elective operations with a 4% incidence. All the other fields were highly objective (incidence of discrepancy rates of 2% or less). Owing to poor reproducibility of the hypertension variable, it was decided to exclude this from the model.

Validation of the tool on an independent dataset showed the model to be a useful predictor of mortality, with an area under the ROC curve of 0.73. This should be compared with an area under the ROC curve of 0.68 for the Parsonnet score for the same period. Based on this initial validation we are happy to use this tool for comparing performance regionally; however, the model will continue to be validated as more data are collected.

Table 5 Per cent mortality by operation for the UK 1997/98, the north west, and individual centres, with 95% CI

Operation	UK Cardiac Register	Region	Centre 1	Centre 2	Centre 3	Centre 4
CABG	2.6 (2.4 to 2.8)	2.6 (2.3 to 3.1)	2.8 (1.9 to 3.9)	2.7 (2.1 to 3.5)	1.82 (1.4 to 2.9)	3.0 (2.3 to 4.0)
AVR	4.5 (3.8 to 5.3)	4.1 (2.8 to 6.0)	2.1 (0.4 to 7.0)	4.4 (2.3 to 8.0)	5.4 (2.0 to 12.7)	4.1 (2.0 to 7.9)
MVR	6.3 (5.3 to 7.6)	4.1 (2.4 to 7.0)	6.4 (1.7 to 18.6)	4.9 (2.2 to 10.2)	2.1 (0.1 to 12.7)	2.9 (0.8 to 8.9)
AVR + CABG	6.4 (5.3 to 7.7)	7.3 (5.1 to 10.2)	7.7 (3.2 to 16.6)	7.1 (4.0 to 12.0)	6.8 (3.0 to 14.0)	8.0 (3.5 to 16.2)
MVR + CABG	12.5 (10.1 to 15.4)	16.1 (10.9 to 23.1)	16.1 (6.1 to 34.5)	19.2 (11.2 to 30.4)	8.7 (1.5 to 29.5)	14.3 (4.7 to 33.6)
DVR +/– CABG	10.5 (8.2 to 13.3)	10.8 (6.6 to 16.9)	14.8 (4.9 to 34.6)	14.0 (6.3 to 27.4)	6.8 (1.8 to 19.7)	8.1 (2.1 to 23.0)

AVR, aortic valve replacement; CABG, coronary artery bypass graft; DVR, double valve replacement; MVR, mitral valve replacement.

Table 6 Multivariate analysis

Variable	*Odds ratio*	*Parsonnet score*	*Per cent discrepancies*
Age	1.06	> 7	0
Renal failure	3.6	10	1
Ejection fraction < 50%	1.4	> 2	8
Hypertension	1.4	3	11
Acute state preoperatively (cardiogenic shock, or IABP not inserted prophylactically)	2.7	10–50	0
Ventilated preoperatively	6.1	0	0
Redo surgery	1.8	> 5	0
Operation priority	1.6	0	4
CABG and mitral valve surgery	4.8	7	0
Double valve surgery	3.4	> 10	0

CABG, coronary artery bypass graft; IABP, intra-aortic balloon pump.

CUSUM CURVES

The risk stratified cusum curves for the four units are given in fig 1. All the curves are similar and for none of the centres does the predicted mortality lie outside the 95% CI of observed mortality.

Discussion

Collecting risk stratified mortality data is important in cardiac surgery, and we have shown that this is possible on all patients undergoing surgery within a defined geographical region in a multicentre study. The northwest region accounts for about one eighth of all the patients undergoing cardiac surgery in the UK. This project was supported in part by supradistrict audit funding, costing approximately £10/patient. The mortality following surgery in the region was 3.5%, which was similar to that reported for the UK in 1997/98.[4] There was no significant difference in crude mortality between the four centres. This suggests that the quality of surgery across the north west is acceptable, which should be reassuring to the surgeons, the hospitals, the purchasers, and the patients in the region.

Comparisons have been made against the cardiac surgical register for 1997/1998,[4] using similar inclusion and exclusion criteria but different definitions of mortality. The cardiac surgical register at this time used 30 day mortality; however, to ensure accurate data collection in this study, the definition of in-hospital mortality given previously was used. In this study patients were classified as alive if they were discharged alive but died within 30 days of the operation; however, patients who remained in any hospital but died after 30 days were classified as dead. Legitimate comparisons were made between the four centres in this study, but some caution should be exercised when comparing these figures with those of the UK cardiac surgical register, owing to these differing definitions. It should also be acknowledged that the mortality returns for the register are not externally validated. It is hoped that in the future, with the introduction of the National Health Service (NHS) number, all mortality—either in or out of hospital—will be easily traceable, albeit for a small cost.

We have collected preoperative risk factors according to agreed definitions and found differences in the incidence of many of these factors across the four centres. Some of these were real differences, such as the incidence of redo surgery. Seven per cent of patients at centre 3 underwent redo surgery compared with 3–4% at the other centres. Redo surgery is known to carry an increased risk and so this is an important finding. However, some of the other

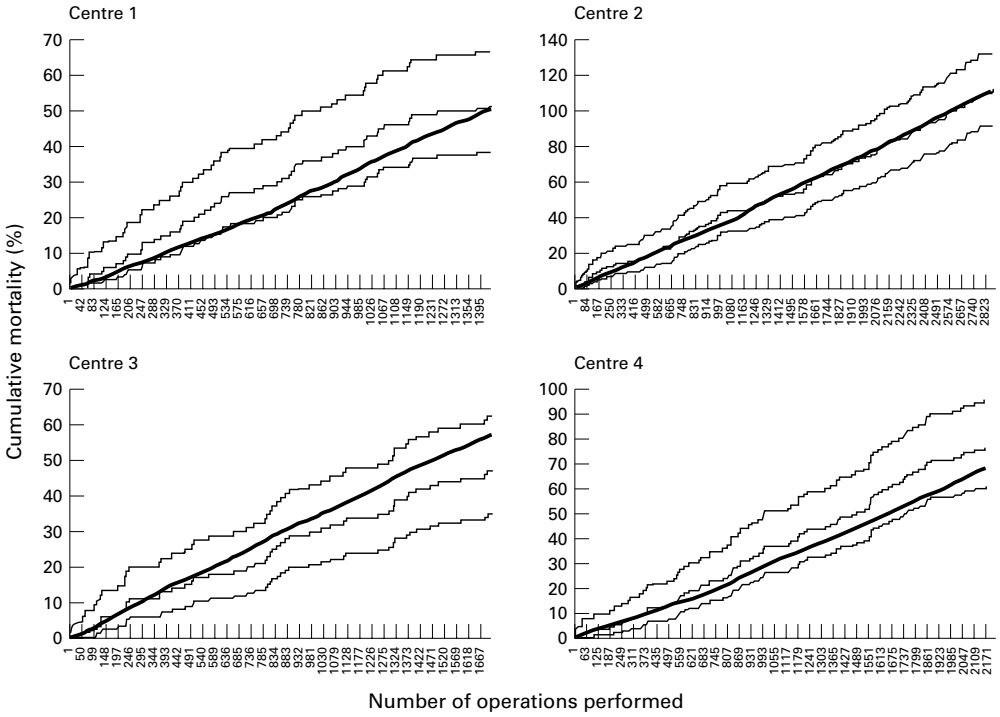

Figure 1 Cumulative mortality with 95% confidence interval (CI) and cumulative predicted mortality by centre. The two thin outlying lines represent 95% CIs around cumulative observed mortality (central thin line), and the thick line represents cumulative predictive mortality working to 0.51 of the Parsonnet score. Performances deviating significantly can be seen by predicted mortality lying either consistently above (better than expected performance) or below (worse than expected performance) the 95% CIs of observed mortality. This was not the case for any of the centres.

differences in incidence were initially less easy to understand. For example, the incidence of respiratory disease was 13%, 9%, 11%, and 29% of cases at centres 1, 3, 4, and 2, respectively. This large difference can be explained on the basis of the definition of respiratory disease (appendix) which is "anyone on drug treatment for respiratory disease or anyone with an FEV_1 < 70% of predicted." Centre 2 is the only one of the four centres that routinely measures FEV_1 (forced expiratory volume in one second), so the true incidence of respiratory disease has been underestimated in the other three centres. Another field with large differences in incidence was ejection fraction < 50%, which was seen in 24%, 34%, 35%, and 45% of cases at centres 4, 1, 3, and 2, respectively ($p < 0.0001$). A poor ejection fraction is usually caused by previous myocardial infarction, which had a steady incidence across the four centres of between 47% and 51% ($p = 0.2$). The apparently low incidence of an ejection fraction < 50% seen at centre 4 probably reflects a different technique for measuring it in the cardiology laboratory. As ejection fraction is one of the items in the Parsonnet score, this would lead to an underprediction of mortality for this centre. These artificial and confounding differences in predicted mortality should be considered when analysing these types of risk stratified mortality data.

Validation checks on the randomly selected cohort of case notes showed that there were some differences in submitted and validated data, leading to a change in Parsonnet score in 19% of cases. Some fields within the Parsonnet score seem to be highly subjective, and including such fields in a predictive model lays the system open to "gaming"—the systematic overprediction of risk resulting from subtle bias in observer recordings of the items which determine the score. These observations should again lead to caution when using the Parsonnet score to cite differences in predicted mortality or observed to expected mortality ratios between centres or individual surgeons.

It is well known that the Parsonnet score includes some factors that are not associated with mortality,[16] and omits some that are, such as urgency of surgery, a risk factor which most surgeons intuitively feel to be important and one which comes out as a significant predictor of operative mortality in our multivariate analysis. The Parsonnet score has also been criticised for its overestimation of risk. For renal failure, age > 70 years, ejection fraction < 50, and double valve replacement, the odds ratios seen from our multivariate analysis are much less than the Parsonnet score weightings, explaining one of the reasons for this.

The risk stratified data have produced some findings that were not expected. There was a significant difference in the predicted risk of the patients operated on in the four centres, and while some caution should be exercised when examining these figures, for the reasons discussed above, the range of Parsonnet scores from 6.2 to 7.6 corresponds to differences in predicted risk of 3.2% and 3.9% between the centres operating on the lowest and the highest risk group of patients. These differences are substantial, considering the close geographical proximity of the four centres and the similar socioeconomic makeup of the populations they serve. It seems likely that the differences will become even greater when centres from other parts of the UK are studied,[6] and it is also likely that there will be differences in predicted risk between the individual surgeons who operate within the four centres as a result of clinical expertise, subspecialisation, and referral patterns. This should lead to serious reservations about using mortality data from centres or individuals for audit or other purposes unless some account has been taken for risk stratification.

In spite of the limitations described above, the Parsonnet score has been shown in this study to have the ability to predict observed mortality in adult cardiac surgery within the useful range. Nevertheless, as it overpredicts risk, it is easy to obtain false reassurance about the quality of surgery. Because one performs better than the Parsonnet score certainly does not mean that one is performing as well as one's peers. Our regional findings of an observed mortality of 0.51 times the Parsonnet score is the first time this type of risk stratified audit standard has been produced from a UK multicentre study, and should be useful for other institutions or surgeons. A recent single centre study[17] has reported an observed mortality of 0.74 times the Parsonnet score; however, there were differences in the exclusion criteria and the definition of mortality used. A further confounding variable in making these comparisons is that the Parsonnet score has been customised since its original publication, and has been changed in different ways by the different centres that have been enthusiastic about its use.[6 17] Consequently, even when Parsonnet risk stratified data are compared between different units, one is not always comparing like with like. For the purposes of this study we have excluded various categories of patients, as described in Methods, notably those undergoing surgery for ischaemic mitral valve disease, ischaemic ventricular septal defects, and left ventricular and aortic aneurysms. These patients together only comprise a small proportion of the total population, but are a group with a high mortality. We feel that it is reasonable to exclude these groups for two reasons. First, patients within each of these groups are heterogeneous, ranging from low to exceedingly high risk, which cannot easily be catered for within the scoring systems. Second, there seems to a differing philosophy between surgeons and units, both for accepting patients with high risk conditions and also for performing left ventricular aneurysm surgery in patients undergoing coronary artery surgery. These two factors combine to muddy the water with respect to risk stratified quality control, which is the primary aim of this type of analysis.

Several different scoring systems have been reported and all have some limitations. Some reservations have been expressed previously about the Parsonnet score, including sugges-

tions from recent studies that its ability to predict mortality is only moderate, that some of the risk factors are subjective, and that many of the items included in the score are not significantly associated with mortality.[6 12] However, the score has strengths in that it is widely accepted by the UK cardiac surgical community, it is easy to use without computer assistance, and it has a useful overall predictive ability for our population.

The cusum plots shown in fig 1 give a simple visual display of the quality of surgery in the four centres. If a centre is performing to a satisfactory standard, the 95%CI of observed mortality should always overlap the predicted mortality. Any trends or temporary periods of good or bad performance can easily be seen. In this study the 95% CI for observed mortality for all centres overlaps predicted mortality at all times.

The use of risk stratified mortality studies for analysing surgical results is obviously a developing area. There is a need for further validation studies on a larger scale than we have performed, to clarify potential errors in datasets such as this, and to see how such errors can influence the results of this type of analysis. It may be that changing techniques of data collection, modifying definitions, or omitting various subjective risk factors from the predictive models will be necessary to obtain robust conclusions and eliminate the potential for gaming. Understanding the ability of the predictive model is important, but of greater importance is the way the models can be used to define the limits of acceptable practice, and be incorporated into programmes to improve standards.[20]

CONCLUSIONS

In summary, we have shown that it is possible to collect risk stratified mortality on all patients undergoing surgery in a defined geographical region at minimal cost. We have seen significant differences in the incidence of many risk factors across the region, and have shown that several of these risk factors are quite subjective. The Parsonnet score is a reasonable overall predictor but observed mortality is 51% of predicted, which should be a useful benchmark for other units or surgeons. There were differences in the predicted mortality between the four centres in the region, but all centres were performing as expected, working to 0.51 times the Parsonnet score.

While the Parsonnet score is a useful predictor, it is subjective, does not contain several risk factors associated with mortality, and does contain some that are not. The score has also been customised in different ways by different centres. The deficiencies demonstrated here should be considered when using Parsonnet risk adjusted mortality to draw conclusions about individual or institutional performance.

We would like to acknowledge to the cooperation given to us by all the cardiac surgeons in the region: Mr Au, Mr Bridgewater, Mr Campbell, Mr Chalmers, Mr Deiraniya, Mr Dhimis, Mr Drakeley, Mr Duncan, Mr Fabri, Mr Fagan, Miss Griffiths, Mr Grotte, Mr Hasan, Mr Hooper, Mr Jones, Mr Keenan, Mr Lawson, Mr Millner, Mr Odom, Mr Page, Mr Rahman, Mr Rashid, and Mr Weir. We would also like to thank the considerable efforts of all the audit teams involved in collecting the enormous amount of data required to carry out such a project. This work has been reported on behalf of the Northwest Regional Cardiac Surgery Audit Group, whose membership is: Mr Au, Mr Bridgewater, Mr Fabri, Dr Jackson, Mr Jones, Mr Keenan, Mrs Silcock, and Mrs Wynne-Jones. This work was supported in part by a Supradistrict audit grant from the Greater Manchester and Merseyside Purchasers.

Appendix: data fields for Northwest Region cardiac surgery audit

Hospital No (for identification and validation)

Date of birth

Sex

Angina status—Canadian Cardiovascular Society classification

Dyspnoea status—New York Heart Association classification

Symptom status:
- Stable: Controlled on drug treatment
- Unstable: Angina requiring admission to hospital and treatment with intravenous medication; previous Q wave myocardial infarction
- A transmural myocardial infarct represented by new Q waves in two or more contiguous leads on ECG

Respiratory:
- No: No history of pulmonary disease
- Chronic obstructive airways disease/emphysema/asthma: patient requires drug treatment for chronic pulmonary disease or FEV_1 less than 75% of predicted value

Cerebrovascular:
- No: No history or symptoms of cerebral or vascular disease
- Yes: Any cerebral neurological deficit including both cerebrovascular accident and transient ischaemic attacks, or previous cerebral surgery

Peripheral vascular disease:
- No
- Yes: History or evidence of aneurysm or occlusive peripheral vascular disease

Extent of vessel disease:
- Normal/single vessel/double vessel/triple vessel/left main stem > 50%

For mitral valve surgery, systolic pulmonary artery pressure

For aortic valve replacement, valve gradient

Ejection fraction:
- Good (> 50%)/fair (30–50%)/poor (< 30%)

Intravenous nitrates:
- No/until operation/within one week of surgery

Last Q wave myocardial infarction:
- Not applicable/< 6 hours/6–24 hours/1–30 days/> 1 month

Previous cardiological intervention:
- Any form of thrombolytic treatment administered within 24 hours of surgery
- Previous PTCA with or without stent
- Previous valvuloplasty

Recent failed intervention:
- Any failed intervention (coronary or valvar) necessitating immediate surgery or surgery in the same admission
- No/failed operation within 24 hours/failed operation in same admission

Previous cardiac surgical intervention

Previous peripheral vascular intervention

Diabetes:
- Any history of diabetes regardless of duration or treatment
- No/oral therapy/diet/insulin

Hypercholesterolaemia:
- A history of fasting cholesterol > 6.5 mmol/l or lower if on treatment

Hypertension:
- A history of blood pressure > 140/90 mm Hg or lower if treated

Smoking:
- Never smoked/still smoking/ex-smoker
- Anybody who has smoked within one month of surgery should be considered to be a current smoker

Renal:
- No: No history of renal disease and creatinine < 200 μmol/l on admission
- Yes: Creatinine > 200 μmol/l or functioning renal transplant, irrespective of creatinine, or on any form of dialysis

Cardiogenic shock:
- Hypoperfusion with a systolic blood pressure < 80 mm Hg and central filling pressure > 20 mm Hg without inotropes, or a cardiac index < 1.8 l/min/m^2, or inotropes + intra-aortic balloon pump required to maintain CI ⩾ 1.8 l/min/m^2 at time of surgery

Ventilated preoperatively

Operation priority:
- Elective: Routine admission from the waiting list
- Urgent: Patients who have not been scheduled for routine admission from the waiting list but who require surgery on the current admission for medical reasons. They cannot be sent home without surgery
- Emergency: Unscheduled patients with ongoing refractory cardiac compromise. There should be no delay in surgical intervention irrespective of the time of day
- Salvage: Patients requiring cardiopulmonary resuscitation en route to theatre, or following the induction of anaesthesia

Operation sequence:
- 1st operation/2nd operation/3rd operation/4th operation/5th operation/6th operation

Operation date

Mortality:
- Death within the same hospital admission as the operation, regardless of cause

Status at discharge:
- Alive/dead

1 Department of Health. *The new NHS*. London: DHSS, December 1997.
2 Treasure T. Lessons from the Bristol case [letter]. *BMJ* 1998;**316**:1685–6.
3 Keogh B, Dussek J, Watson D, *et al*. Public confidence and cardiac surgical outcome [letter]. *BMJ* 1998;**316**:1759–60.
4 Society of Cardiothoracic Surgeons of Great Britain and Ireland. *Cardiac surgery register 1997/8.*
5 DuBois RW, Rogers WH, Moxley JH, *et al*. Hospital inpatients mortality: is it a predictor of quality. *N Engl J Med* 1987;**317**:1674–80.
6 Bridgewater BJM, Neve H, Moat N, *et al*. Predicting operative risk for coronary artery surgery in the united kingdom; a comparison of various risk prediction algorithms. *Heart* 1998;**79**:350–5.
7 Parsonnet V, Dean D, Bernstein, AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart disease. *Circulation* 1989;**701**(suppl): 13–112.
8 Nashef SAM, Carey F, Silcock MM, *et al*. Risk stratification for open heart surgery: trial of the Parsonnet system in a British hospital. *BMJ* 1992;**305**:1066–7.
9 Allen A, Forsyth AT. Risk stratification in open heart surgery [letter]. *BMJ* 1992;**305**:1500–1.
10 Spiegelhalter DJ. Risk stratification for open heart surgery [letter]. *BMJ* 1992;**305**:1500.
11 Treasure T. Risks and results of surgery. *Br Heart J* 1995;**74**: 11–12.
12 Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;**240**:1285–93.
13 O'Connor GGT, Plume SK, Olmstead EM, *et al*. Multivariate prediction of in hospital mortality associated with coronary artery bypass graft surgery. *Circulation* 1992;**85**: 2110–18.
14 Dupuis J-Y, Wynands JE. Risk-adjusted mortality to assess quality of care in cardiac surgery. *Can J Anaesth* 1993;**40**:91–7.
15 Marshall G, Shroyer LW, Grover FL, *et al*. Bayesian-logit model for risk assessment in coronary artery bypass grafting. *Ann Thorac Surg* 1994;**57**:1492–500.
16 Gabrielle F, Roques F, Michel P, *et al*. Is the Parsonnet's score a good predictive score of mortality in adult cardiac surgery? Assessment by a French multicentre study. *Eur J Cardiothoracic Surg* 1997;**11**:400–11.
17 Poloniecki J, Valencia O, Littlejohns P. Cumulative risk adjusted mortality charts for detecting changes in death rates: observational study of heart surgery. *BMJ* 1998;**316**: 1697–700.
18 Hamilton L. *Statistics with STATA 5*. London: Duxley Press, 1998.
19 de Leval M. Analysis of a cluster of surgical failures. *J Thorac Cardiovasc Surg* 1994;**107**:914–23.
20 O'Connor G, Plume SK, Olmstead EM, *et al*. A regional intervention to improve the hospital mortality associated with coronary artery bypass graft surgery. *JAMA* 1996;**275**: 841–6.